



Published in final edited form as:

J Biomed Inform. 2013 June ; 46(3): 444–457. doi:10.1016/j.jbi.2013.02.003.

A method for estimating from thermometer sales the incidence of diseases that are symptomatically similar to influenza

Ricardo Villamarín*, Gregory Cooper, Michael Wagner, Fu-Chiang Tsui, and Jeremy U. Espino

Center for Advanced Study of Informatics in Public Health, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Early detection and accurate characterization of disease outbreaks are important tasks of public health. Infectious diseases that present symptomatically like influenza (SLI), including influenza itself, constitute an important class of diseases that are monitored by public-health epidemiologists. Monitoring emergency department (ED) visits for presentations of SLI could provide an early indication of the presence, extent, and dynamics of such disease in the population.

We investigated the use of daily over-the-counter thermometer-sales data to estimate daily ED SLI counts in Allegheny County (AC), Pennsylvania. We found that a simple linear model fits the data well in predicting daily ED SLI counts from daily counts of thermometer sales in AC. These results raise the possibility that this model could be applied, perhaps with adaptation, in other regions of the country, where commonly thermometer sales data are available, but daily ED SLI counts are not.

Keywords

Thermometers; Symptomatically-like influenza; Estimation; H1N1

1. Introduction

The 2009 H1N1 influenza outbreak highlighted the need for methods that not only detect an outbreak but also estimate incidence so that public-health decision makers can allocate appropriate resources in response to such an outbreak. Several studies [1–3] have attempted to estimate incidence during the 2009 H1N1 outbreak, but they were performed months after the outbreak and were based on data that are either not easy to collect (e.g., confirmed laboratory cases) or not available in a timely fashion (e.g., surveys). Some influenza surveillance systems were also in operation during the H1N1 outbreak, but they did not provide daily estimates of influenza activity [4,5].

* Corresponding author. Address: 5607 Baum Boulevard, 5th floor, Pittsburgh, PA, 15206-37041, USA. ricardo.villamarin@alumni.pitt.edu (R. Villamarín).

Previous research indicates that data on over-the-counter (OTC¹) sales of retail healthcare products, such as cough and cold medications, can support the early detection of disease outbreaks [6,7]. In this paper, we provide support that thermometer sales (TS) increase significantly and early during an influenza outbreak.

We explore the hypothesis that TS data can be used to estimate accurately the number of patient cases that present to EDs and are symptomatically like influenza (SLI). We developed and evaluated a model to estimate the number of SLI cases in EDs in Allegheny County (AC), Pennsylvania based on TS. To train this model, we used ED visit data from hospitals in AC and TS data from AC. We then generalized the model to predict daily ED SLI cases from the daily TS in any region in which we are able to collect TS data. Unlike counts of ED SLI cases, which are not readily available for most counties in the US, there exists TS data for a significant number of locations throughout the US, with at most a one day delay. Given this availability, the method and model we present is intended to help public health epidemiologists throughout the US to estimate the extent of an influenza outbreak as it presents to EDs with little delay; the evaluation of this method beyond AC is proposed as a next step in the future pursuit of this line of research.

The remainder of this paper is organized as follows. In the next section, we provide an overview of previous research that aimed to detect and characterize influenza outbreaks using OTC sales. We summarize a project that facilitates obtaining daily OTC sales data across the US with little delay. We then discuss a method we developed to estimate SLI cases in AC. We describe the experimental methodologies we used to evaluate the performance of the models in AC, and we present and discuss the results of those experiments. We present methods to adjust TS data so that these models can produce estimates for other regions, not just AC. Finally, we present conclusions and close the paper with a summary.

2. Background and related work

This section provides background information about the use of OTC healthcare-product sales to detect disease outbreaks. We then provide an overview of the 2009 H1N1 influenza epidemic in the US and of several published studies that estimate its level of activity. The section then describes a system we have developed to estimate from ED reports the expected number of daily SLI cases that visit EDs in a region. We close the section by describing two influenza surveillance systems that can use the predictions produced by the models we present in this paper.

2.1. OTC sales data

Several studies have measured the value of using OTC-related data for the detection of outbreaks. These are either survey-based studies that are designed to reveal healthcare seeking behavior or retrospective studies to understand the correlation between OTC data and outbreaks.

¹All the abbreviations used in this paper are listed in a glossary located after the References section.

Survey-based studies have shown that a high percentage of individuals prefer to self treat using OTC healthcare products when they feel ill before, and sometimes instead of, seeking medical attention [8,9,7]. Metzger et al. [7] found that among the individuals they surveyed, purchasing an OTC healthcare product was the most common first action (37% of the time) to treat their illnesses.

Retrospective studies show high correlation between known outbreaks and the sales of OTC healthcare products. Campbell et al. [10] found that for some OTC categories the correlation is high enough that they could detect seasonal outbreaks of respiratory diseases based on OTC sales earlier than methods that use physician office visits (up to 7 days earlier). Magruder [11] obtained similar results in a study of the National Capital area (Washington DC and its suburbs), and found that, on average, sales of influenza remedies preceded office visits by 2.8 days and that there was an average correlation of 0.90 between them. Das et al. [12] found that the ratio of sales of cough and influenza medicines to sales of analgesics was highly correlated with the ratio ED SLI visits to other syndrome ED visits, with a correlation coefficient (r^2) of 0.6. In this previous research, the OTC data used were collected only from the individual locations (mostly cities) that were studied. In the next section we describe a service that collects OTC data from across the US.

2.2. The NRDM project

The National Retail Data Monitor (NRDM) project has greatly reduced the effort to obtain OTC sales data for use in biosurveillance. NRDM collects daily sales data for products in 18 OTC categories across the country [13,6]. Public health departments can access such information through a web-based interface or via web service protocols. NRDM was developed in our lab, the Real-time Outbreak and Disease Surveillance Laboratory (RODS), with the cooperation of selected retail chains that sell OTC medications. Using data from NRDM, our laboratory studied the effects of a single outbreak and other health events relative to the sales of OTC products. One of these internal studies, for instance, found that TS increased significantly and early during the 2003–2004 influenza outbreak [6]. The models we present in the current paper use TS from NRDM to estimate SLI cases that present to EDs and in the population.

Although the NRDM is a powerful information resource, it has several limitations. First, in most areas of the US, there are retail stores selling OTC healthcare products that do not report data to the NRDM. Second, for many areas where OTC sales data are collected, the OTC *market share* of some of the retail store chains that provide data to NRDM is not known with certainty. Third, the daily sales number reported for a product category is the sum of units that are sold *and the units that are returned*. Thus, a negative number of reported units on a particular date means that there were more units of a product category returned (which were bought at earlier, unknown dates) than units sold (which can be zero) on that date. Likewise, a positive number of units reported simply means that there were more units sold than units returned (which can be zero) on that day. Also, a zero number of units reported means that there was the same number of units sold and returned on that date, whereas when nothing is reported from a store for a given product type (i.e., no sales record is created), it means that no units of that type were sold or returned.

We have addressed these limitations in multiple ways. In areas where there is incomplete coverage and market share information is available, we make adjustments to the sales data. When market share information is lacking, we estimate it. Finally, if we aggregate sales data over a large enough region (such as a county), the effects of returned units are not as significant. These methods are described in Appendix A.

2.3. H1N1 2009 outbreak

In April of 2009, pandemic H1N1/09 influenza infected thousands in Mexico and the US. Although influenza activity during the summer of 2009 in the US was low, starting in late August of that year a second wave of influenza caused by H1N1 occurred across the country [14,15]. It peaked during the second half of October, and finally it largely subsided in December [14,15,3,16]. Although it is difficult to determine exact dates, the literature indicates a period from August 23rd, 2009 (the start of the “last two weeks of August” [14]) through December 12th, 2009 (“declined to less than 10% during the week ending December 12” [15]). In AC, the epidemic had its highest activity level in late October [3].

Due to difficulties in obtaining an accurate count of total H1N1 cases [2], several researchers developed methods to estimate those counts and other characteristics of the outbreak. Early attempts to estimate the incidence and severity of pandemic H1N1 in the US include studies by Reed et al. [1] and Presanis et al. [2]. The former used a multiplicative model coupled with a Monte Carlo simulation to predict that between 1.8 and 5.7 million H1N1 cases had occurred in the US from April to July 2009. The latter used a Bayesian evidence-synthesis framework to estimate that during the same period as Reed et al. between 0.16% and 1.44% of symptomatic H1N1 patients were hospitalized and that between 0.0007% and 0.048% died. Later in 2009, the Centers for Disease Control and Prevention (CDC) used the method described in Reed et al. to estimate that between April and December 12, 2009 around 55 million people (approximately 18% of the US 2009 population) had been infected with H1N1 in the US [16].

Ross et al. conjectured that asymptomatic and mild cases were missed by the CDC estimates [3]. To partially overcome that limitation, they performed a study to determine the seroprevalence of antibodies to the pandemic 2009 H1N1 influenza strain in residents of AC. They found that seroprevalence against pandemic H1N1 across age groups was approximately 21%. They then extrapolated the results from their sample to both the local county population and the general US population. They estimated that about 63 million people in the entire US became infected through early December 2009.

Table 1 shows a summary of the methods described in this section.

2.4. Bayesian Case Detection (BCD) system

Tsui et al. have developed the Bayesian Case Detection (BCD) system that provides a differential diagnosis of a patient case that is obtained from an electronic medical record. Its current implementation can diagnose SLI diseases and shigellosis [17]. The version of BCD we used takes as input free-text ED reports. The free-text ED reports are collected from several EDs in AC. BCD has a natural language processing component that extracts patient symptoms and signs from a free-text ED report and codes the clinical findings with Concept

Unique Identifiers (e.g., “the patient reports a severe cough today” would yield Cough = present) from the Unified Medical Language System (UMLS). These coded findings are input to a Bayesian network model of disease, which outputs a posterior probability over the modeled diseases that the patient might have.

Tsui et al. performed a preliminary evaluation of the BCD system to detect SLI cases using free-text reports from seven EDs in AC [17]. Laboratory-confirmed positive and negative cases of influenza were used as gold standards. The researchers obtained an area under the ROC curve of 0.80 (95% CI: 0.76–0.85). We used the BCD system to estimate the number of daily SLI cases that presented to eight monitored EDs in AC. These estimates are computed by summing the posterior SLI probabilities of the patients who visited those EDs on a given date.

A limitation of the BCD system is that it only monitored approximately 39% of all EDs in AC during 2009, due to the limited number of hospitals providing ED reports to it. We refer to that percentage as BCD's *ED coverage*. Since we know this ED coverage, we are able to estimate the total number of ED cases in all of the AC EDs. We explain this adjustment in Section 3.

2.5. Complementary approaches

This section describes two surveillance systems that utilize methods different than the one we introduce. It also discusses how the method we introduce can complement those systems.

The CDC's outpatient influenza-like-illness (ILI) surveillance network (ILINet) collects information on patient visits to more than 3000 healthcare providers for influenza-like illness in all 50 states, the District of Columbia and the US Virgin Islands. ILINet defines ILI as “fever (*temperature of 100°F [37.8 °C] or greater*) and a cough and/or a sore throat in the absence of a KNOWN cause other than influenza” [5]. Thus, in this particular context, ILI is a type of SLI that considers only those three specific symptoms. ILINet reports on a weekly basis ILI activity for the entire US, as well as individually for each of the ten major US surveillance regions that the CDC has defined, usually with a 1–2-week reporting lag [4]. The CDC also uses data reports collected through ILINet to create a measure of ILI activity by state. The limitations of ILINet are that (1) the information it provides has a 1–2 week reporting lag, and (2) the finest geographical granularity for its estimates is the state level, although the CDC only broadly shares the larger, region-level numerical data.

The models presented in this paper complement ILINet. ILINet reports ILI activity at many types of outpatient clinics [5] at a regional level with a 1–2 week lag. The methods we presented here provide ED-derived estimates at the county level with just a one-day lag.

Google, Inc. developed “Flu Trends” [4] to predict ILI activity across the US based on Internet searches made using the Google search engine. To build Flu Trends, Google researchers performed a variable selection experiment to isolate the top search terms, which they call the “ILI-related search queries.” Using the fraction represented by those queries as the explanatory variable, they created a linear model to predict the weekly ILI percentages across the nine major US regions reported by the CDC [5]. They found that the predictions

made by the resulting model were highly correlated to the CDC weekly estimates of ILI [1] over the US at the regional level. They also found that they could consistently predict ILI percentages one to two weeks prior to the publication of the CDC weekly reports.

Despite its usefulness, the Google Flu Trends model has limitations. First, the high correlations observed are only validated for large populations (they were computed between estimates and data at the CDC's surveillance region level, and at the state level for Utah). Second, like the ILINet estimates, Flu Trends estimates are for weekly activity. Although in principle Flu Trends estimates could be produced daily (i.e., using queries made during the last day), to our knowledge the accuracy of predictions for this shorter time period are not available and have not been reported.

The models in this research also complement Flu Trends. Whereas the latter makes *weekly* predictions at the regional and state levels based on Internet searches, the current research can provide *daily* estimates at the county level.

3. A method for estimating SLI cases from TS

In this section we present a novel method for estimating SLI cases from OTC TS in EDs. Due to data availability, we created this model for estimating ED SLI cases in AC, but the methodology is flexible and can be extended to make estimates for other regions (Section 6).

To create a model for AC, we first gathered training data as follows. We obtained non-negative daily sales of pediatric and adult thermometers from May 1 through December 31, 2009 ($n = 245$ days) in AC from the NRDM. We will refer to the data from this period as the 2009 *dataset*. We compared these TS data with the expected number of daily SLI cases that presented to EDs monitored in AC during the same period according to BCD. Fig. 1 shows a plot of TS during that eight-month period (solid line) and the number of ED SLI cases estimated by BCD during the same period (dash-dotted line). We found that the strongest cross-correlation of the two curves is 0.9 when the expected cases lag behind the TS data by one day; a zero-day lag had almost the same cross-correlation (0.89). Fig. 2 shows a plot of the estimated cases as a function of TS, which supports a linear relationship between those two quantities (black line).

The rationale for the selection of the 2009 dataset is twofold. First, the BCD system officially started operating in AC in May 2009, and hence, we only had data starting at that time. Second, we wanted to devise a model that could estimate from thermometer sales not only SLI activity during outbreaks (high activity periods) but also during non-outbreak periods (low activity periods). The 2009 dataset contains sub-periods of both high (during the 2009 H1N1 outbreak) and low SLI activity and thus was the suitable for creating the aforementioned model. A model that is capable of estimating SLI activity only during outbreaks is more limited in its usefulness.

Eq. (1) describes a linear model of SLI as a function of TS for AC, where S and I denote the slope and intercept parameters of the model, respectively, ED_SLI_BCD represents SLI cases in the EDs in AC that have BCD deployed, and TS_NRDM denotes thermometer sales reported by NRDM for AC.

$$ED_SLI_BCD(AC, date) = I + S \cdot TS_NRDM(AC, date) \quad (1)$$

Eq. (1) only predicts the number of SLI cases that present to EDs in AC that have BCD installed. To predict the total number of ED cases of those diseases in AC, we need to adjust that equation to incorporate the ED coverage of BCD in AC. Let $ED_SLI_T(AC, date)$ be the total number of ED SLI cases (the T stands for total) during $date$ in AC. We can estimate ED_SLI_T using Eq. (2), where $EDcov$ is the ED coverage of BCD, which is discussed in more detail later in this section.

$$ED_SLI_T(AC, date) = ED_SLI_BCD(AC) / EDcov(AC) \quad (2)$$

We combine Eqs. (1) and (2) to obtain Eq. (3) that approximates the total number of daily SLI cases that present to all EDs in AC.

$$\begin{aligned} ED_SLI_T(AC, date) &= \frac{I + S \cdot TS_NRDM(AC, date)}{EDcov(AC)} \\ &= \frac{1}{EDcov(AC)} + \frac{S}{EDcov(AC)} \cdot TS_NRDM(AC, date) \end{aligned} \quad (3)$$

At the time this research was performed, the ED coverage of BCD in AC was around 39%. We calculated it by dividing the number of visits to EDs in AC where BCD is deployed by the total number of visits to EDs in AC from May 1st, 2009 through December 31st, 2009. We obtained the latter count from a system that our laboratory has deployed in all EDs in AC, except for one ED in a federal hospital.

For the sake of example, suppose that in the model given by Eq. (3) the slope (S) is 0.5 and the intercept (I) is 20. If on a given day d , the net number of thermometers sold in AC is 10, then the model would predict that the total ED SLI cases in AC on that day is approximately $64 = 20/0.39 + (0.5/0.39) \cdot 10$, where 0.39 is the estimated ED coverage of BCD in AC.

4. Evaluation methodology

This section describes experiments that evaluate how well Eq. (1) performs when using AC data for training a model with data from one period of time of m days and testing with data from a subsequent period of n days. To do so, we create training and testing sets, perform least squares linear regression on the training set to obtain slope (S) and intercept (I) parameters, and then use Eq. (1) with those parameters to predict ED cases for each day in the test set. The next training period begin n days after the start of the former training period.

For example, for a training set of $m = 42$ days (6 weeks) and a test set of $n = 14$ days (2 weeks) we proceed as follows. We train a model with TS and BCD data from for example 5/6/2009 (a Sunday) through 6/16/2009; we then apply that model to predict BCD-estimated SLI cases from 6/17/2009 through 6/30/2009. The TS data and BCD estimates from 5/20/2009 (i.e., $n = 14$ days after 5/6/2009) through 7/02/2009 (i.e., 42 days after 5/20/2009) become the next training set.

To evaluate the performance of the models after the 2009 H1N1 outbreak had largely dissipated, we included the first three months of 2010. For the sake of uniformity, we start on 5/6/2009, which is a Sunday, instead of on 5/1/2009, and include data up to 4/3/2010, which is a Saturday. We will refer to this dataset of 333 days as the *experimental dataset*. A plot of that set is shown in Fig. B1 in Appendix A.

Based on the above arrangement, we performed three experiments with different training and testing set sizes. In all these experiments the smallest training set consists of three weeks of data, and larger training sets increase in periods of three weeks, with the largest set consisting of 15 weeks. That is, $m = 21, 42, 63, 84$, and 105 days. We stopped at 105 days because that is approximately the number of days in the primary wave of the 2009 H1N1 outbreak in AC. The experiments differ in the number of days used for the testing set:

- Experiment 1 (E1): Each testing set consists of 7 days (1 week) of data. The total number of training sets in this experiment is 190.
- Experiment 2 (E2): All the testing sets consist of 14 days (2 weeks) of data. The total number of training sets is 94.
- Experiment 3 (E3): Each testing set consists of the same number of days as its corresponding training set. That is, $n = m$. The total number of training sets is 28.

The rationale of picking the sizes of the testing sets in the first and second experiment is to see how well the model can estimate disease activity during periods of time close to the date when the estimate is produced, which may be of greater public-health interest. The reason for selecting the size of the testing set on the third experiment was to see how well the models performed on a testing set the same size as its training set.

We evaluated the goodness of the regression performed to create each of the linear models described above with the *coefficient of determination*, denoted as R^2 . It measures the fraction of the variability of the modeled response variable (e.g., expected number of daily ED SLI cases computed by BCD) that is explained by a regression [18]. Relationships with a strong positive (or negative) slope that are close to being linear would be expected to have an R^2 value near 1. Associations with a slope near zero would be expected to have an R^2 value near zero. We calculate R^2 with Eq. (4) [18], where

- m is the number of days in the *training* set used to generate a model (i.e., m can be 21, 42, 63, 84, and 105).
- $PredictedValue_i$ represents a given model's prediction of daily ED SLI cases from thermometer sales on day i of the training set,
- the variable $ActualValue_i$ is the number of SLI cases estimated by BCD on that same day, and
- the bar ($\bar{}$) denotes the arithmetic mean of the quantity below it (during the given period of n days).

$$R^2 = 1 - \frac{\sum_{i=1}^m (ActualValue_i - PredictedValue_i)^2}{\sum_{i=1}^m (ActualValue_i - \overline{ActualValues})^2} \quad (4)$$

We evaluated the predictive performance of each of those models using a measure known as the *mean absolute percentage error* (MAPE) [19,20], which is computed using Eq. (5). A MAPE value of X means that, on average, a model's predictions are $X\%$ greater or smaller than the actual quantity. Thus, the closer the MAPE values are to zero (with zero indicating that the predictions and actual values are the same), the better is the forecasting performance of an estimation method. As in the case of R^2 , the value of n in Eq. (5) for this experiment is the number of days in the *testing* set for which we generate a prediction (i.e., n can be 7, 14, 21, 42, 63, 84, and 105), the variable $ActualValue_i$ in Eq. (5) is the number of SLI cases estimated by BCD on day i in that set, the variable $PredictedValue_i$ represents a given model's prediction of daily ED SLI cases from thermometer sales on that same day, and the parallel bars (\parallel) denote the absolute value of the quantity they enclose.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|ActualValue_i - PredictedValue_i|}{ActualValue_i} \cdot 100 \right) \quad (5)$$

To assess the forecasting performance of a method, Lewis [21] suggested four ranges of MAPE values, namely, [0%, 10%], (10%, 20%], (20%, 50%], and (50%, ∞). He labeled these ranges as having highly accurate, good, reasonable, and inaccurate forecasting performance, respectively. We use these as regions of reference to facilitate our analysis.

In the next section we summarize the performance for each experiment using the average MAPE, the minimum and maximum MAPE values, and the minimum and maximum R^2 values.

5. Results and discussion

In this section we present and discuss the results of the experiments outlined in Section 4. These experiments evaluate models that estimate daily ED SLI cases in AC during different periods of time. The result of experiments E1, E2, and E3 are shown in Tables 2–4, respectively. Table 2 shows that experiment E1 achieves the lowest Min MAPE values, relative to the other experiments, and achieves average MAPE values of 16.7% or less. However, E1 also has the largest MAPE estimation errors, as given by the Max MAPE column in Table 2). Recall from Section 4 that a MAPE greater than 10% and less than or equal to 20% is considered to be good forecasting performance, whereas one greater than 50% is inaccurate performance [21]. Table 2 shows that the average MAPE is well within the good forecasting range. A more stringent criteria, however, is to consider the maximum MAPE over all 44 sets. When we do so, most (4 out of 5) of the maximum MAPE values in E1 are uncomfortably close to inaccurate performance. As a result, the models produced by the method presented in this paper to estimate SLI cases over a period as short as 7 days should be used with caution.

Table 3 shows that the average MAPE values are almost as good as those of E1 (the training sets are of the same sizes in both experiments but the testing sets are of different sizes, except for the first), but the maximum MAPE values are not as high. All of the former are in the “good” forecasting bracket, whereas all the latter are well within the “reasonable” forecasting bracket.

For experiment E3, Table 4 shows that the average MAPE values are not as good as those obtained in E1 and E2, but the maximum MAPE values are lower than in the other two experiments. It is also clear that in this experiment the errors for the last two configurations ($m = n = 84$, and $m = n = 105$) are higher than those of the other configurations for this same experiment. The latter seems to occur when the training set contains mostly data from a non-outbreak period but the majority of data of the testing set comes from an outbreak period (or vice versa).

We now briefly discuss individual results in experiments E1 and E2, rather than their aggregated statistics. Figs. 3 and 4 show, for experiment E2 and E3, respectively, the MAPE values for each testing set in the configuration where the training set consists of $m = 42$ days (6 weeks) of data and the testing set consists of $i = 14$ days of data. Appendix B includes similar figures for the other configurations in these experiments.² In these figures it can be seen that the forecasting performance of the models declines under two circumstances:

- A decline occurs when there is a sharp drop in the number of SLI cases in a test set relative to the associated training set. For example, the testing set consisting of data from 10/29/2009 through 11/11/2009 ($n = 14$ days) shown in Fig. B2, is a period when the number of BCD-estimated SLI cases was falling sharply. However, its corresponding training set consisting of data from 10/7/2009 through 10/28/2009 ($m = 21$ days) coincides with the period of highest disease activity. A similar phenomenon happens for the latter (Fig. B2) and other configurations, for the testing sets that start at various points during December 2009 (e.g., Figs. 3 and B3), and during February 2010 (e.g., Figs. B4 and B6).
- The second case is the reverse of the first, namely, the training set includes mostly days of low disease activity whereas the testing set includes days of noticeably higher activity. The latter circumstance is most clearly visible for configurations consisting of training sets with data from up to July and August, 2009, but with testing sets with data from September 2009 and afterward. For example, see Fig. B7 (second bar) for $m = n = 63$, and Fig. B5 (second bar) for $m = 105$, $n = 14$.

The performance decline in the two cases described above can be explained in terms of a mismatch between the training and testing periods, due to changes in disease activity between those periods. The forecast errors worsened both before periods of high disease activity and after periods in which the disease activity had peaked and was waning. Fortunately, such performance declines did not last for more than a single estimation period in the majority of cases. For example, this pattern is observed for the testing periods that

²Except for the configurations where $m = n = 83$ and $m = n = 105$ in E3. The only two MAPE values resulting from that experiment came from the first and second testing sets on each of those two configurations. They are shown in the Max MAPE and Min MAPE columns in Table 4, respectively.

start on 7/15/2009 and on 12/16/2009 in Fig. 3, and those that start on 9/9/2009 and on 12/02/2009 in Fig. 4. In both figures it can be seen that each of those periods is followed by an estimation period with a noticeably lower estimation error. The same transient decline is visible in the other figures as well. One possible approach to addressing this problem would be to combine the methods presented in the present study with outbreak detection techniques. The detection of the start of an outbreak and the beginning of its decay would then guide which data to use to train the models.

In Tables 2–4 the Max R^2 values were consistently associated with periods of high and dynamic disease activity, where we would expect strong positive relationships between TS and daily ED SLI cases. In contrast, Min R^2 values were generally associated with periods of no disease activity, where the slope of the relationship between TS and ED SLI cases is close to zero.

We close this section with a summary of the main findings from the experiments described here.

- On the whole, the linear models achieved “good” forecasting (MAPE) performance under most circumstances, and with only a few exceptions they achieved “reasonable” forecasting performance otherwise.
- The lowest forecasting errors (as measured with the MAPE metric) are obtained when estimating 7 days of data, but such a short period also produces the worse forecasting errors. However, by using prediction periods of 14, 42, and 63 days the worst-case forecasting performance improves while the average forecasting error remains in the interval of “good” performance.
- Forecasting performance generally decreases when a model is trained primarily with data from a period of high disease activity but used to forecast a period of mostly low disease activity, or vice versa. However, such declines in performance are transient.

6. Extensions

In this section, we show how to extend the model developed in Section 4 to predict daily ED SLI cases in other US second-level administrative divisions (SLDs), such as counties and independent cities [22,23]. For the sake of simplicity, we will use the term “county” instead of the more precise term SLD. The main assumption behind this method is that the linear relationship we found for AC between daily TS and ED cases of SLI (Fig. 2) is also present in other counties. We believe this is a reasonable first-order approximation. However, due to the non-availability to us of ED reports from other counties, we do not attempt here to validate the accuracy of the AC model when applied to other counties, but leave it as future research.

Eq. (3) is only applicable to predict cases in AC's EDs based on TS in AC, as reported by NRDM. We extended the model by adjusting the AC model for differences between (1) the population size of AC and a given county, and (2) the TS in AC and that county. To do so, we follow a three-step procedure. First, we normalize thermometers sold to be per person

(PP) in the total population (pop) of a county. We use NRDM's market share (MS) in that county to estimate the TS. NRDM's MS (which we denote with MS_NRDM) is the fraction of the total OTC sales by all retailers in a particular county that are reported by NRDM retailers. We explain how we obtain MS_NRDM information in Appendix A. In Eq. (6), TS_PP is thermometer sales per person, r denotes the county of interest, $pop(r)$ represents its population, and d is a particular date.

$$TS_PP(r, d) = \frac{TS_NRDM(r, d) / MS_NRDM(r)}{pop(r)} \quad (6)$$

Second, we incorporate Eq. (6) into Eq. (3) and adjust it to produce an estimate for AC of daily ED SLI cases per person in the population. The result is shown in the following equation.

$$\begin{aligned} ED_SLI_PP(AC, d) &= \frac{\frac{I}{EDcov(AC)} + \frac{S}{EDcov(AC)} \cdot TS_PP(AC, d) \cdot (pop(AC) \cdot MS_NRDM(AC))}{\frac{I}{EDcov(AC) \cdot pop(AC)} + \frac{S \cdot pop(AC)}{EDcov(AC)}} \cdot TS_PP(AC, d) \cdot MS_NRDM(AC) \quad (7) \\ &= \end{aligned}$$

Note that in Eq. (7), the term $MS_NRDM(AC)$ is part of the calculation of the adjusted slope (Eq. (3)), whereas $TS_PP(AC, d)$ serves as the *independent* variable. Hence, to calculate ED SLI cases in other regions, we replace $TS_PP(AC, d)$, which is specialized for AC, with the more general term $TS_PP(r, d)$, which is the thermometer sales per person in an arbitrary country r , as shown in the following equation.

$$ED_SLI_PP(r, d) = \frac{I}{EDcov(AC) \cdot pop(AC)} + \frac{S}{EDcov(AC)} \cdot TS_PP(r, d) \cdot MS_NRDM(AC) \quad (8)$$

To obtain an estimate of the number of daily ED SLI cases in a county, denoted with ED_SLI , we simply adjust ED_SLI_PP by multiplying it by the population of the county. We obtain Eq. (9) as a result. Note that if we use $r = AC$, Eq. (9) reduces to Eq. (3), which is the model in the previous section to predict total ED cases for AC.

$$ED_SLI(r, date) = ED_SLI_PP(r, date) \cdot pop(r) \quad (9)$$

Continuing with the example of the previous section (where for illustration purposes we assumed that the S and I parameters were 0.5 and 20, respectively), suppose that we wish to compute the total ED cases in some county SC in which a net of 30 thermometers were sold on day d by the retailers in SC that report to NRDM. Furthermore, suppose that the population of SC is 200,000 people, that MS_NRDM in SC is 0.6 (i.e., 60% of all TS in SC are made by NRDM retailers), and that MS_NRDM in AC is 0.7. Using the population of AC in 2009 (1,218,494 people [23]), Eq. (9) predicts that there were about 13 ED cases of SLI in SC on day d :

$$ED_SLI(SC, d) = \left(\frac{20}{0.39 \cdot (1,218,494)} + \frac{0.5}{0.39} \cdot \left(\frac{30/0.6}{200,000} \right) \cdot 0.7 \right) \cdot 200,000 = 12.9 \quad (10)$$

7. Conclusions and future work

Based on the results of the experiments presented here, we conclude that it is possible that linear models trained to estimate SLI ED cases based on thermometer sales can have good forecasting performance. We showed this in the case for Allegheny County (AC), Pennsylvania. In our experiments we found that under most testing conditions the methods achieved good forecasting performance, with only occasional, transient declines. Given these results and the timeliness with which the TS data can be obtained from NRDM, we think the methods presented in this paper can be helpful for epidemiologists and public health officials.

The results presented here raise the possibility that similar models could be applied in other regions of the country, where thermometer sales data are commonly available, but daily ED SLI counts are not. In this paper, we describe how to transform the AC model to produce such estimates. At the time that this study was performed, we did not have access to estimates of daily ED SLI cases in other counties in the US. Thus, we could not test how accurately the AC-derived model predicts ED SLI counts in other counties in the US, many of which do have TS data that are available through the NRDM. Even more broadly, it will be important to test how well the approach described here will work in countries other than the US. In addition, in the future, it will be useful to evaluate how well models developed using data from one outbreak are able to perform in making predictions in subsequent outbreaks. We view these tasks as important directions for future research.

We hypothesize that other signals of SLI activity, such as laboratory confirmed SLI cases, may exhibit a similar linear relationship with TS, as well as with the sales of other relevant OTC medications. Investigating such relationships is another promising direction for future research.

Acknowledgments

This project was funded by CDC Grants P01 HK000086 and 1U38 HK000063-01, NIH Grant R01 LM009132, and NSF Grant IIS-0911032.

We thank Mr. Nicholas Millett for his programming contributions to the implementation of our estimation techniques as web services and also for developing the web service we used to create all the maps included in this paper.

Appendix A. NRDM's market share estimation

This section describes a method for estimating the market share (MS) of NRDM in a region, which appears in Eq. (6). This method is based on information provided by a market analysis company.

A.1. Market share report

We obtained a copy of the *Drug Industry Market Share Report* (DIMSR) from Chain Store Guides LLC (CSG), a market analysis company. It contains information about drug retail operators in more than 900 Core Based Statistical Areas (CBSA) in the US [24]. A CBSA is defined by the US. Census Bureau as a metropolitan statistical area (MSA) or a micropolitan area. An MSA is an urban center of 50,000 or more people, while a micropolitan area is an urban center of at least 10,000, but less than 50,000 people [25]. The CBSAs consist of constituent regions, but not all of a given CBSA's constituent regions need to be located in a single state. For example, the Philadelphia CBSA is comprised of counties in several states (see Fig. A1). The most common type of constituent region of a CBSA is the county, but other types of administrative divisions, such as independent cities, municipalities, and districts [22,23], are also possible.

CSG obtains sales information of OTC healthcare products (including, but not limited to, thermometers) of a subset of retailers located in each of the CBSAs that CSG covers. The fraction of OTC sales by a particular retailer in a given CBSA is referred to in the DIMSR as the *front-end market share* (FEMS) of that retailer in that CBSA. As an example, Table A1 shows information for fictitious retailers with stores in a fictitious CBSA labeled *FC* (note that the FEMS of those retailers may be different in another CBSA in which they are also present). Of the five retailers in *FC*, three report sales data to CSG (and thus have an associated FEMS), three report data to NRDM, two report data to both, and one reports data to neither (see Fig. A1 and Table A1).

Because TS are OTC products, it would be the best to use the FEMS to approximate MS_{NRDM} in a region. However, the absence of FEMS for many retailers in the CBSAs prevents us from doing so. Fortunately, DIMSR also provides information about the *prescription market share* (RXMS) of *all* the retailers in all the CBSAs that are part of that report. Thus, in the next section we describe a method for calculating MS_{NRDM} that is based on the RXMS of the retailers that report to NRDM.

A.2. Method

We address the approximation of NRDM market share in an arbitrary region in two-steps using only CBSA data from the DIMSR. First, we decompose each $CBSA_i$ in DIMSR into its constituent regions, such as counties. Second, for each constituent region of $CBSA_i$, we add up the RXMS of each of the retailers in that constituent region that report sales to NRDM. We do this because we want to be able to produce estimates at the county level, not at the CBSA level. Hereafter, we will refer to the approximation produced with this method as MS_{NRDM_RX} . Equation (A.1)³ shows the computation of MS_{NRDM_RX} , where r is a particular constituent region (e.g., AC) and $Retailer_{NRDM}$ denotes a retailer that reports OTC sales data to the NRDM and that is present in the CBSA of interest.

³We will denote an approximation of a quantity that we cannot directly measure with a hat (^) over such symbol for that quantity. For instance, $MS_{\hat{NRDM}}$ means an approximation of MS_{NRDM} .

$$\begin{aligned}
 MS_ \hat{NRDM}(r) &= MS_ NRDM_ RX(r) \\
 &= \sum_{Retailer_{NRDM}} RXMS(Retailer_{NRDM}, r) \quad (A.1)
 \end{aligned}$$

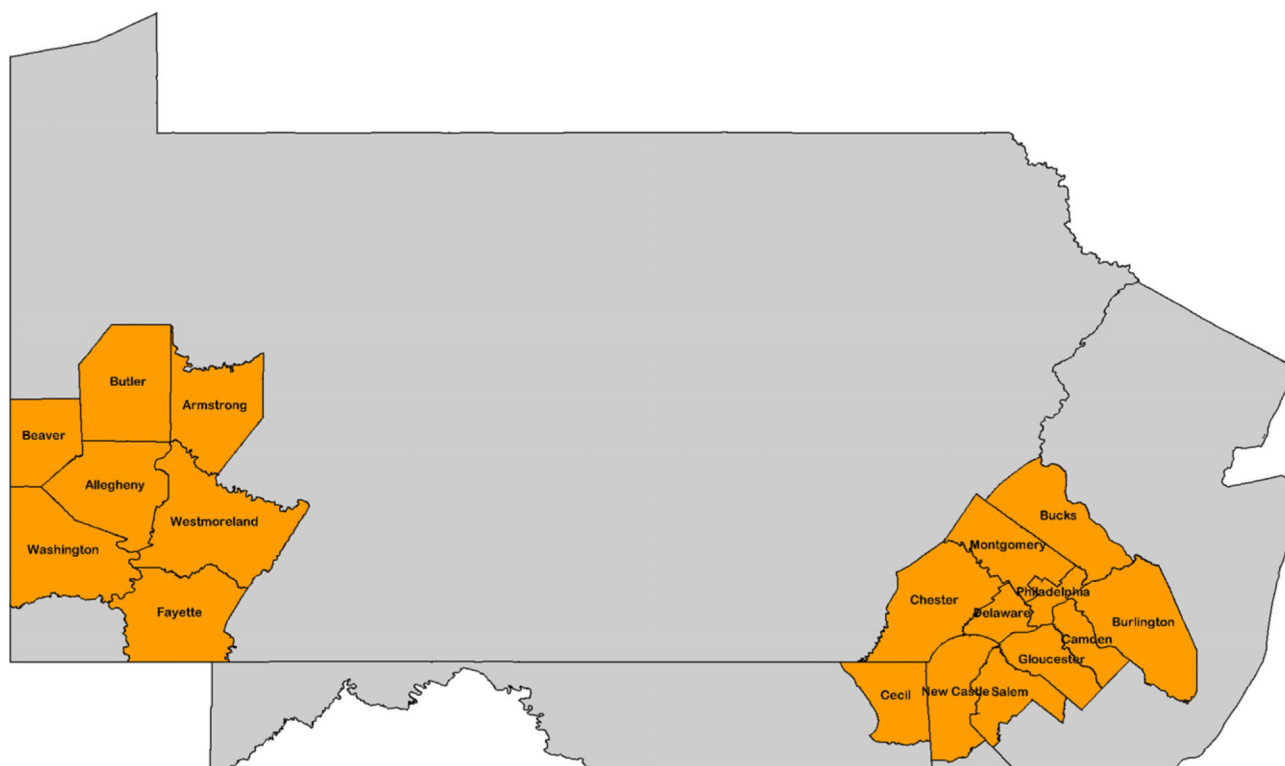
This method makes two main assumptions. First, it assumes that each of the retailers in a given CBSA has the same RXMS in each constituent region in that CBSA. For example, if the RXMS of a given retailer in the Pittsburgh CBSA (Fig. A1) is 7%, we assume that its RXMS is also 7% in each of the counties (e.g., Allegheny, and Butler) that are considered part of that CBSA. While no doubt imperfect, we believe this is a plausible first-order approximation, because retailers often have a commercial penetration that is regional. Second, we assume that the RXMS of any retailer in a CBSA is roughly equivalent to its real FEMS of OTC healthcare products in that CBSA. That is, if the sales data that CSG uses for calculating FEMS were available from *all* the retailers in a CBSA, as is the information used to calculate RXMS is, we conjecture that these two quantities would be approximately the same. The more that each individual obtains his or her outpatient prescriptions and OTC products from the same retail pharmacy chain, the better this approximation will be.

We use MS_NRDM_RX as an approximation of MS_NRDM in Equation (6) (from Section 6) as shown below, to calculate the thermometer sales per person, TS_PP , in a given constituent region r .

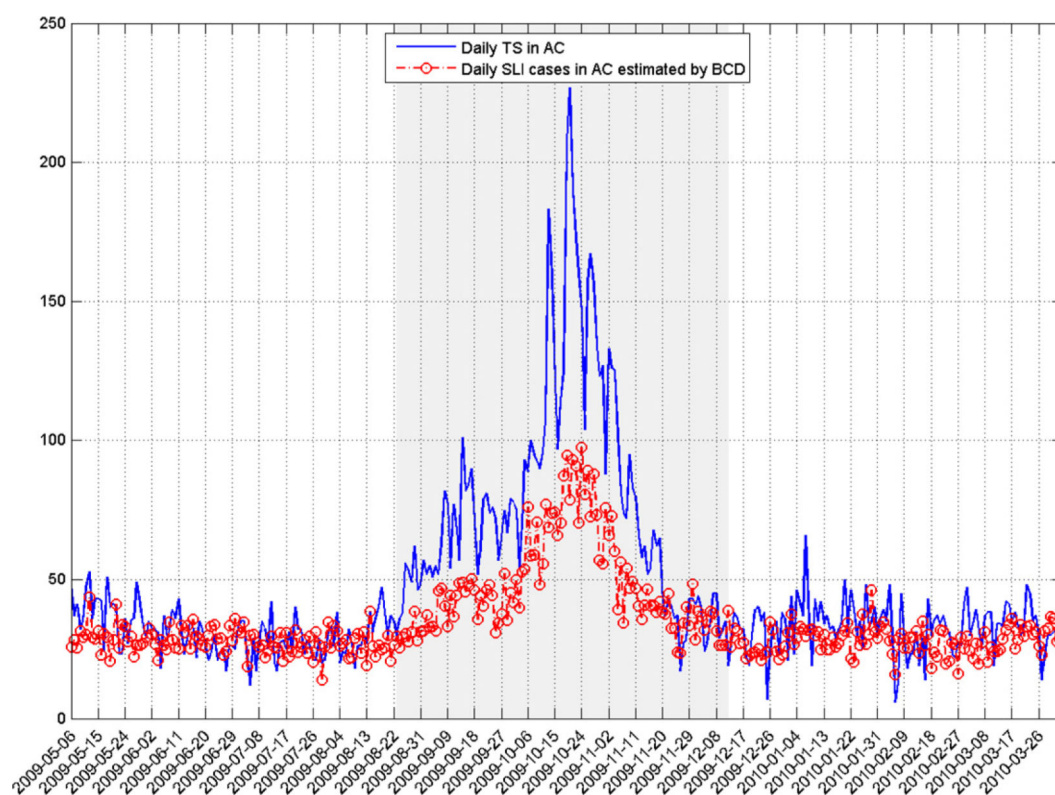
$$TS_PP(r, d) = \frac{TS_NRDM(r, d) / MS_NRDM_RX(r)}{pop(r)} \quad (A. 2)$$

Appendix B

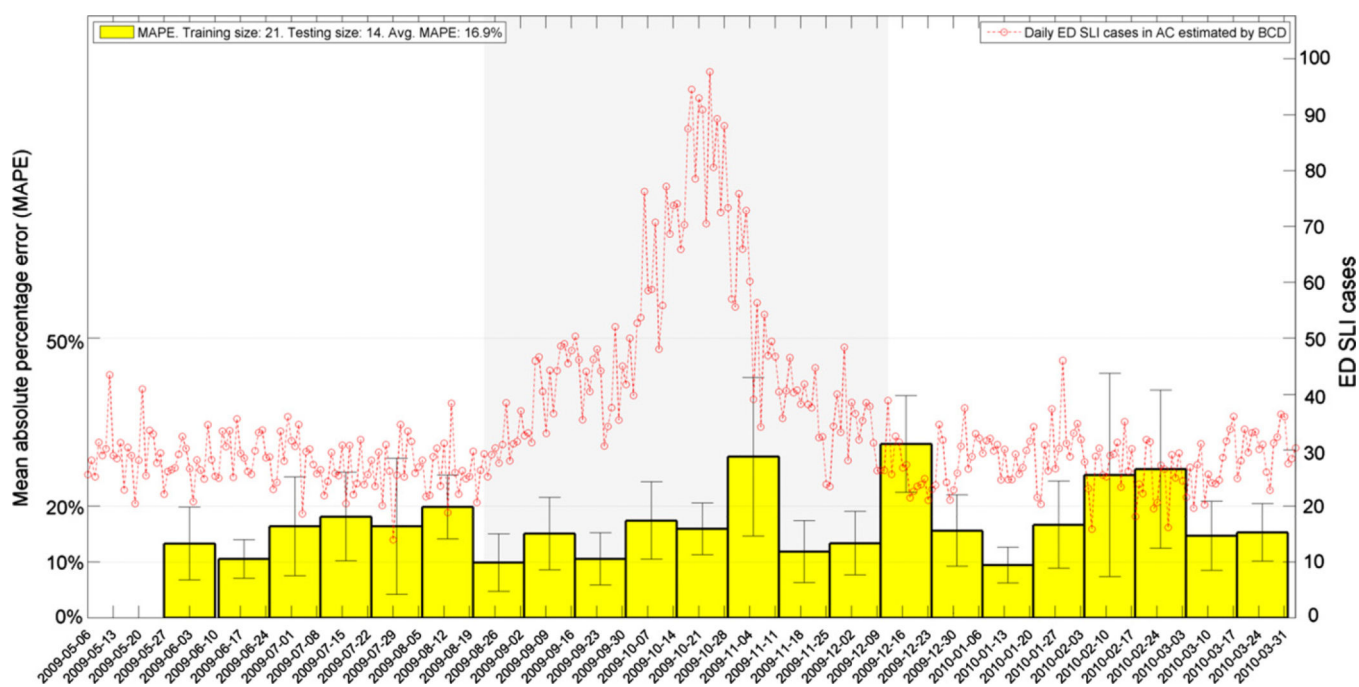
(See Figs. B1–B7)

**Fig. A1.**

Two sample CBSAs. On the left is the Pittsburgh CBSA, which only contains counties located in Pennsylvania (PA), including Allegheny County. On the right is the Philadelphia CBSA, which encompasses regions in PA and other states (e.g., New Jersey).

**Fig. B1.**

Plots of daily TS and daily ED SLI cases estimated in BCD from May 1, 2009 through March 31, 2010 in AC.

**Fig. B2.**

MAPE values (yellow bars, left y axis) for one configuration in experiment E2 when the training set consists of 21 days of data and the testing set consists of 14 days of data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

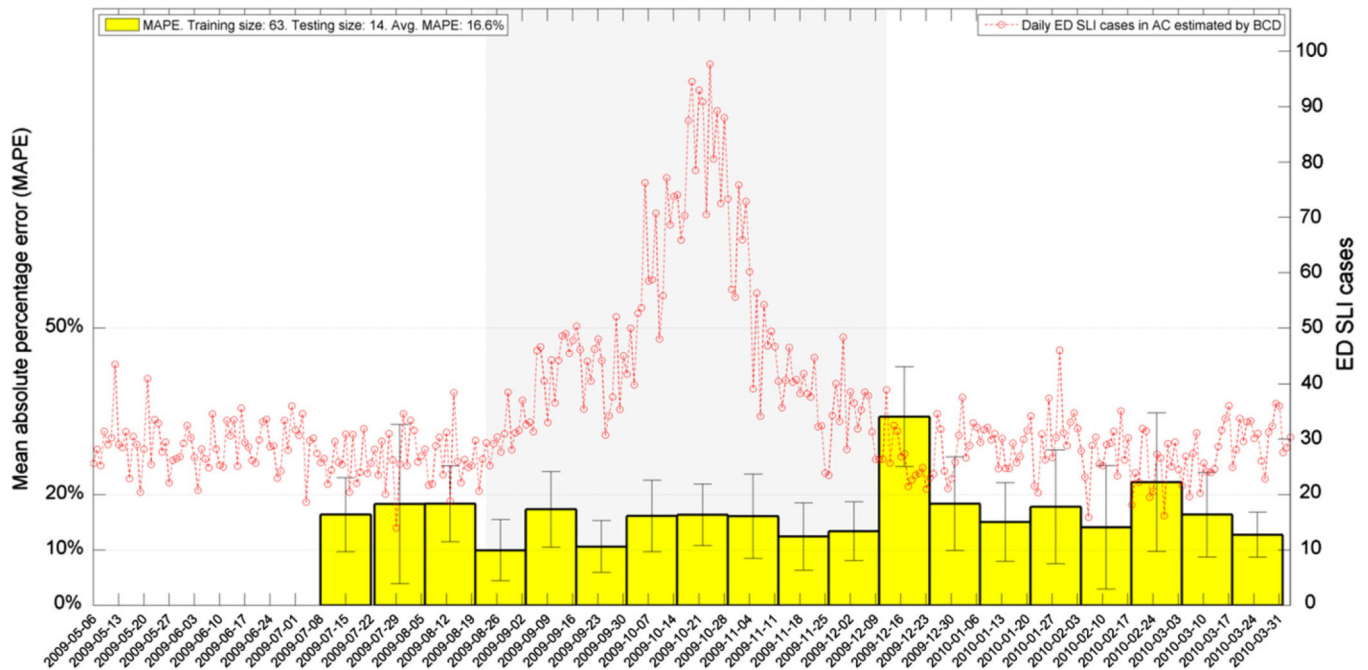
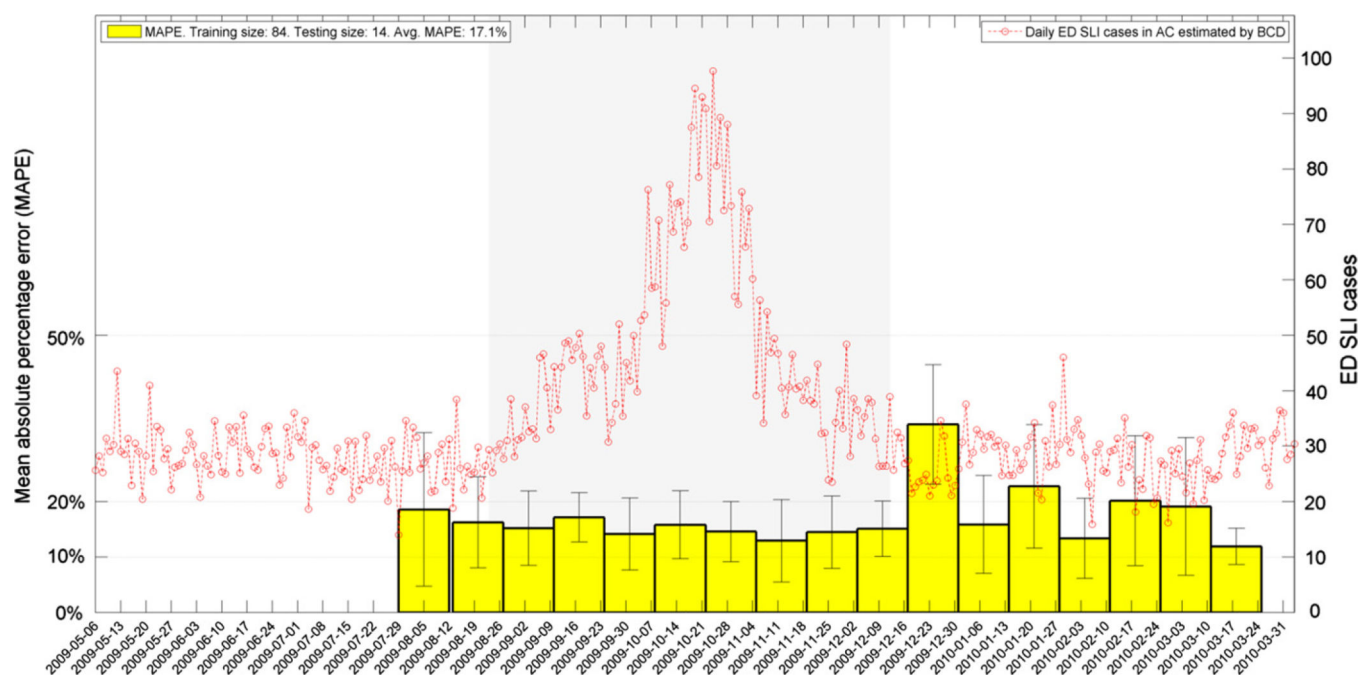
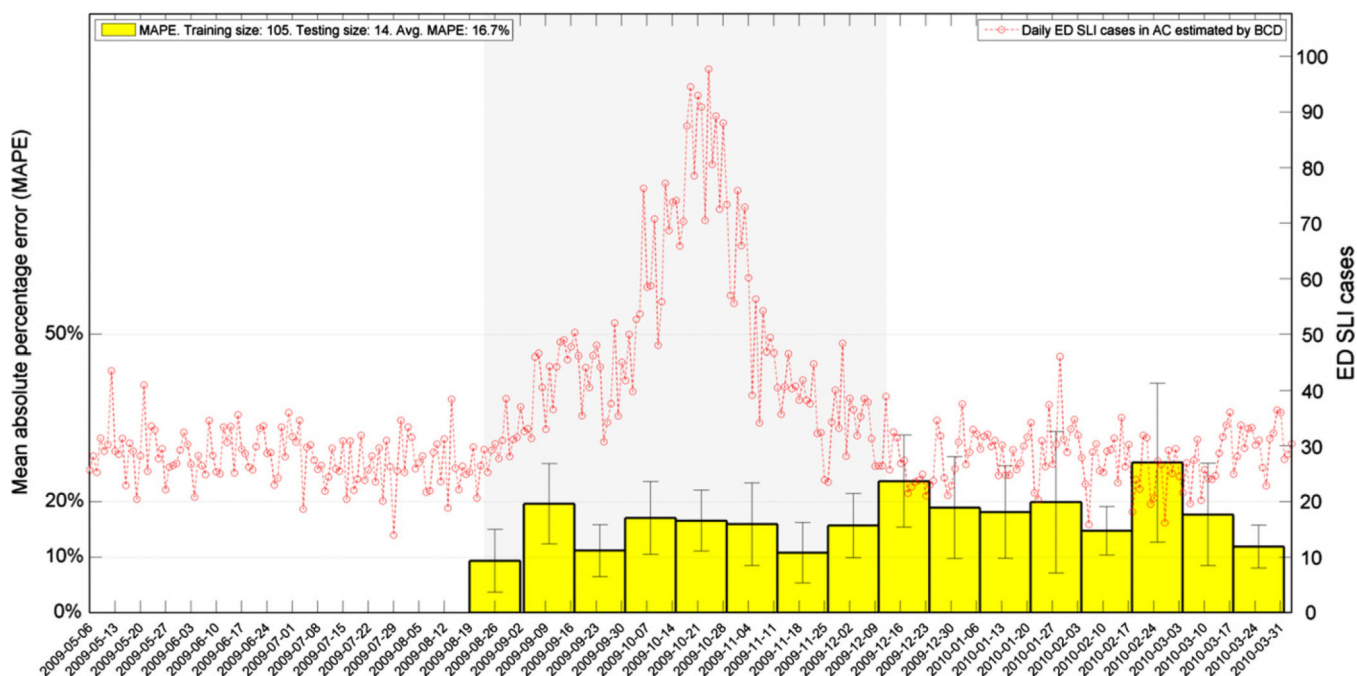


Fig. B3.

MAPE values (yellow bars, left y axis) for one configuration in experiment E2 when the training set consists of 63 days of data and the testing set consists of 14 days of data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. B4.**

MAPE values (yellow bars, left y axis) for one configuration in experiment E2 when the training set consists of 84 days of data and the testing set consists of 14 days of data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. B5.**

MAPE values (yellow bars, left y axis) for one configuration in experiment E2 when the training set consists of 105 days of data and the testing set consists of 14 days of data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

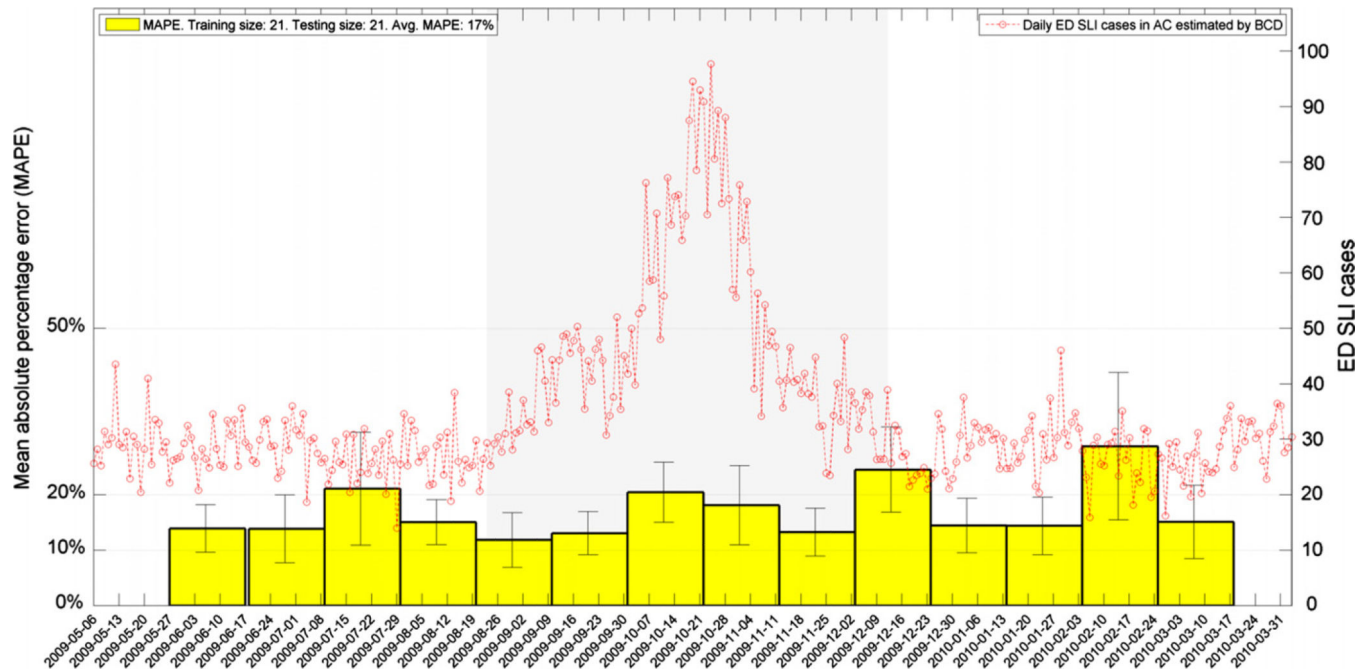
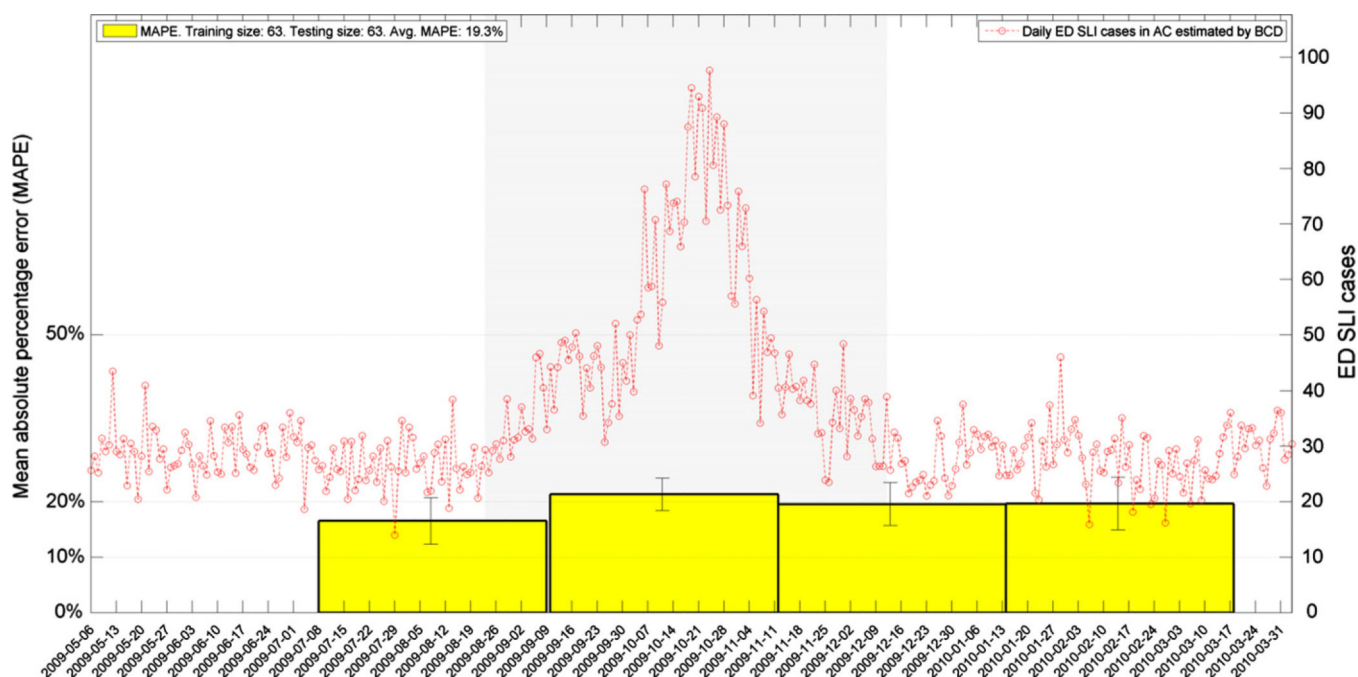


Fig. B6.

MAPE values (yellow bars, left y axis) for one configuration in experiment E3 when the training set consists of 21 days of data and the testing set consists of 21 days of data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. B7.**

MAPE values (yellow bars, left y axis) for one configuration in experiment E3 when the training set consists of 63 days of data and the testing set consists of 63 days of data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table A1

Example of the information that might be reported in the DIMSR for fictitious retailers located in a fictitious CBSA called *FC*.

List of all the retailers in <i>FC</i>	Reports to NRDM?	OTC sales data are in DIMSR?	FEMS of the retailer in <i>FC</i>	RXMS of the retailer in <i>FC</i> (%)
Retailer 1	Yes	Yes	75%	70
Retailer 2	Yes	Yes	10%	9
Retailer 3	Yes	No	Not estimated in DIMSR	6
Retailer 4	No	Yes	15%	10
Retailer 5	No	No	Not estimated in DIMSR	5

Glossary

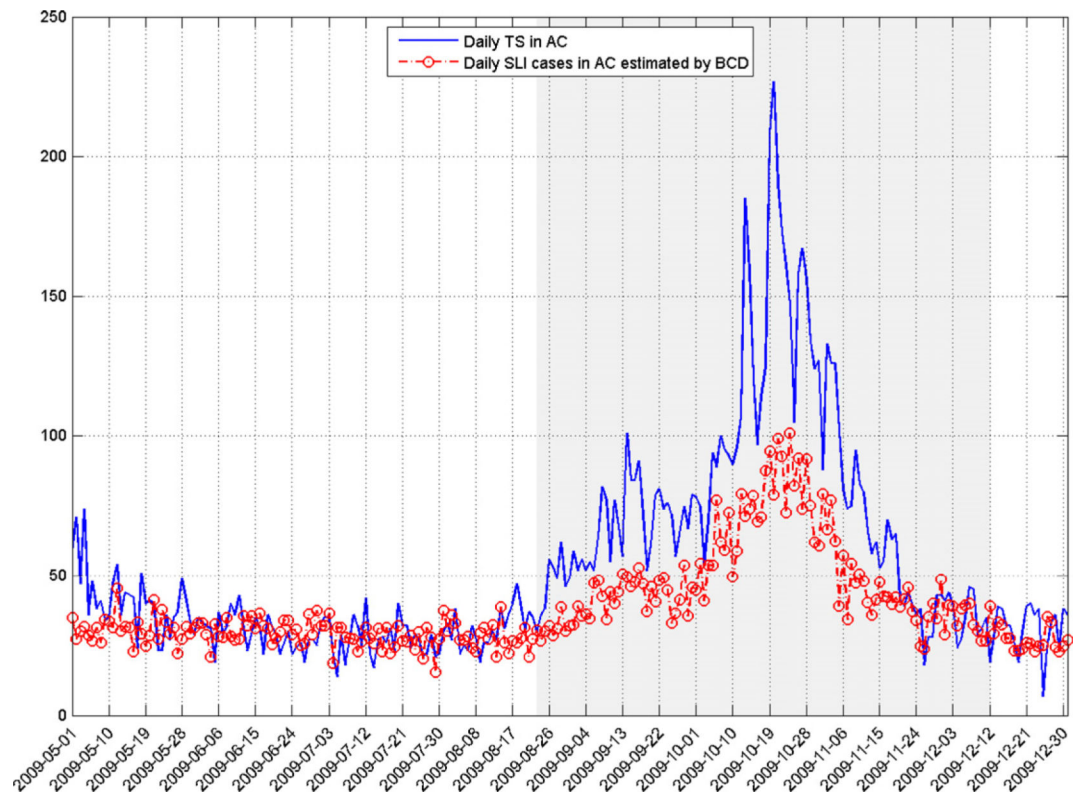
AC	Allegheny County, Pennsylvania
BCD	Bayesian Case Detection system (Section 2.4)
CBSA	core based statistical area (Appendix A)
CDC	centers for disease control and prevention

CI	confidence interval
CSG	chain store guides LLC (Appendix A)
DIMSR	drug industry market share report (Appendix A)
FEMS	front-end market share (Appendix A)
ILI	influenza-like illness (Section 2.5)
MAPE	mean absolute percentage error (Section 4)
MS	market share
MSA	metropolitan statistical area (Appendix A)
NRDM	the National Retail Data Monitor project (Section 2.2)
OTC	over-the-counter sales (Section 2.1)
PA	Pennsylvania
POP	population
PP	per person
ROC	receiver operator characteristic
Rx	a prescription for medicine
RXMS	prescription market share (Appendix A)
SLD	US second-level administrative division (Section 3)
SLI	A disease that presents symptomatically like influenza, including influenza itself
TS	Thermometer sales
UMLS	Unified Medical Language System

References

1. Reed C, Angulo FJ, Swardlow DL, Lipsitch M, Meltzer MI, et al. Estimates of the prevalence of pandemic (H1N1) 2009, United States, April–July 2009. *Emerg Infect Dis.* 2009; 15:2004–7.
2. Presanis AM, De Angelis D, The New York city swine flu investigation team. Hagy A, Reed C, et al. The severity of pandemic H1N1 influenza in the United States. A Bayesian analysis. *PLoS Med.* 2009; 6(12):e1000207. doi:10.1371/journal.pmed.1000207. [PubMed: 19997612]
3. Ross, T.; Zimmer, S.; Burke, D.; Crevar, C.; Carter, D.; Stark, J., et al. [11.11.10] Seroprevalence following the second wave of pandemic 2009 H1N1 influenza.; *PLoS Curr Influenza.* 2010. p. 24 [RRN1148]<URL: <http://www.ncbi.nlm.nih.gov/pubmed/20191082>>
4. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature.* 2008; 457(7232):1012–4. [PubMed: 19020500]
5. Centers for disease control and prevention (CDC). [28.06.11] Seasonal influenza (Flu) – overview of influenza surveillance in the United States. <<http://www.cdc.gov/flu/weekly/overview.htm>>; 2011
6. Hogan, WR.; Wagner, MM. Sales of over-the-counter healthcare products.. In: Wagner, MM.; Moore, AW.; Aryel, RM., editors. *Handbook of biosurveillance.* Academic Press; Amsterdam, Boston: 2006.

7. Metzger KB, Hajat A, Crawford M, et al. How many illnesses does one emergency department visit represent? Using a population-based telephone survey to estimate the syndromic multiplier. *MMWR Morb Mortal Wkly Rep.* 2004; 53(Suppl):106–11. [PubMed: 15714638]
8. Labrie, J. Self-care in the new millenium: American attitudes towards maintaining personal health. Consumer Healthcare Products Association; Washington (DC): 2001.
9. McIsaac WJ, Levine N, Goel V. Visits by adults to family physicians for the common cold. *J Fam Pract.* 1998; 47:366–9. [PubMed: 9834772]
10. Campbell M, Li CS, Aggarwal C, Naphade M, Wu KL, Zhang T. An evaluation of over-the-counter medication sales for syndromic surveillance. In: IEEE international conference on data mining-life sciences data mining workshop. doi: 10.1142/9789812772664_0007.
11. Magruder S. Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of human disease. *Johns Hopkins Univ Appl Phys Lab Tech Digest.* 2003; 24:349–53.
12. Das D, Metzger K, Heffernan R, et al. Monitoring over-the-counter medication sales for early detection of disease outbreaks New York City. *MMWR Morb Mortal Wkly Rep.* 2005; 54(Suppl): 41–6. [PubMed: 16177692]
13. Wagner MM, Robinson JM, Tsui FC, Espino JU, Hogan WR. Design of a national retail data monitor for public health surveillance. *J Am Med Infor Assoc.* 2003; 10(5):409–18. <http://dx.doi.org/10.1197/jamia.M1357>.
14. CDC Novel H1N1 Flu | The 2009 H1N1 Pandemic: Summary Highlights. April 2009–April 2010; <<http://www.cdc.gov/h1n1flu/cdcresponse.htm>> [accessed: 15.02.12].
15. Update: influenza activity – United States. 2009–10 Season. <<http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5929a2.htm>> [accessed 15.02.12].
16. Centers for disease control and prevention (CDC). [28.09.11] CDC estimates of 2009 H1N1 influenza cases, hospitalizations and deaths in the united states. Apr-Dec. 2009 <http://www.cdc.gov/h1n1flu/estimates/April_December_12.htm>
17. Tsui, F-C.; Espino, J.; Sriburadej, T.; Su, H.; Dowling, J. Building an automated Bayesian case detection system.. Conference of the international society for disease surveillance.; Park city, Utah: ISDS; 2010.
18. Jain, R. The art of computer system performance analysis: techniques for experimental design, measurement, simulation, and modeling. John Wiley & Sons; New York (NY): 1991.
19. Makridakis, SG.; Wheelwright, S.; Hyndman, R. Forecasting: methods and applications. John Wiley & Sons Inc.; 1998.
20. Yaffee, RA.; McGee, M. Introduction to time series analysis and forecasting: with applications of SAS and SPSS. Academic Press, Inc.; Orlando, (FL, USA): 2000.
21. Lewis, CD. Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting. Butterworth Scientific; London (UK): 1982.
22. Census Bureau. [30.09.11] Population and population change by CBSA Status. <<http://www.census.gov/popest/metro/tables/2009/CBSA-EST2009-11.csv>>; 2009
23. Census bureau. [30.09.11] Population estimates. <<http://www.census.gov/popest/gallery/maps/maps-county2009.csv>>; 2009
24. Chain Store Guides LLC. Drug Industry Market Share, Report. 2009
25. Census Bureau. [31.03.12] Geographic terms and concepts – core based statistical areas and related statistical areas. <http://census.gov/geo/www/2010census/gtc/gtc_cbsa.html>.

**Fig. 1.**

Plots of daily TS and daily ED SLI cases estimated in BCD from May 1, 2009 through December 31, 2009 in AC. The shaded area indicates the period when the second wave of the 2009 H1N1 outbreak is estimated to have occurred throughout the US (see Section 2.3).

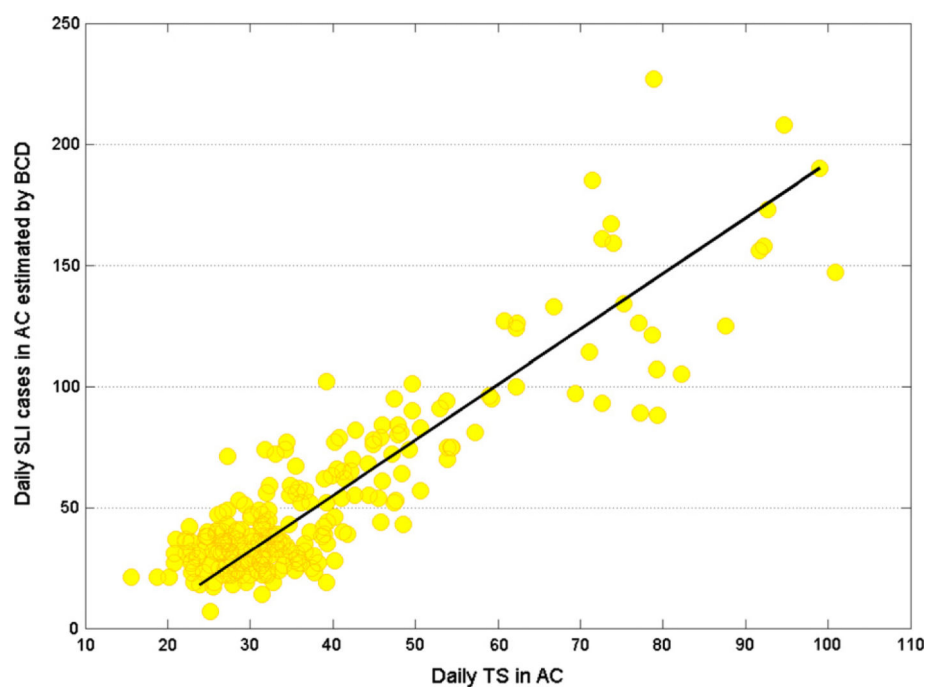


Fig. 2.
A plot of daily estimates of ED SLI cases as function of daily TS in AC from May 1, 2009 through December 31, 2009.

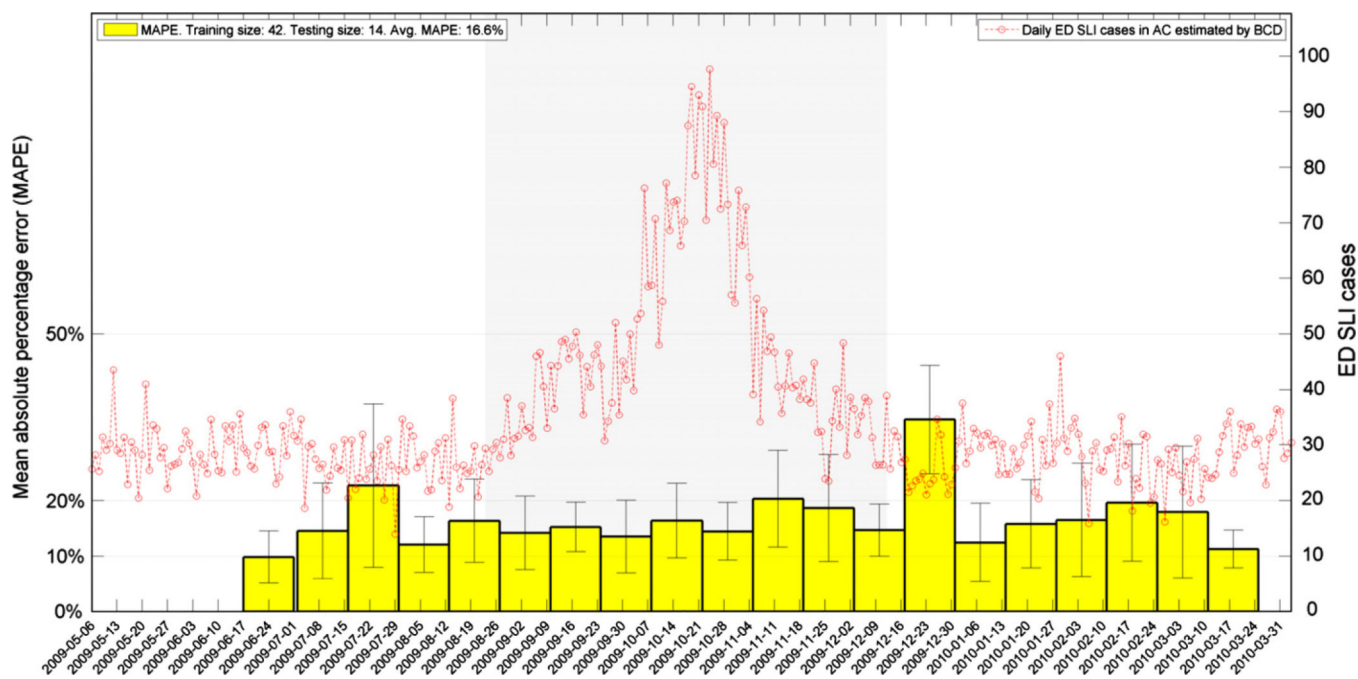


Fig. 3.

MAPE values (yellow bars, left y axis) for one configuration in experiment E2 when the training set consists of 42 days (6 weeks) of data and the testing set consists of 14 days of data. The error bars represent a 95% C.I. The bars span over the testing period from which the MAPE value was computed. The shaded area corresponds to the second wave of the 2009 H1N1 outbreak (Section 2.3). The daily estimates of ED SLI cases by BCD (Section 3) are also shown (red line, right y axis). The tick marks on the left y axis correspond to the MAPE ranges suggested by Lewis [21] (Section 4). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

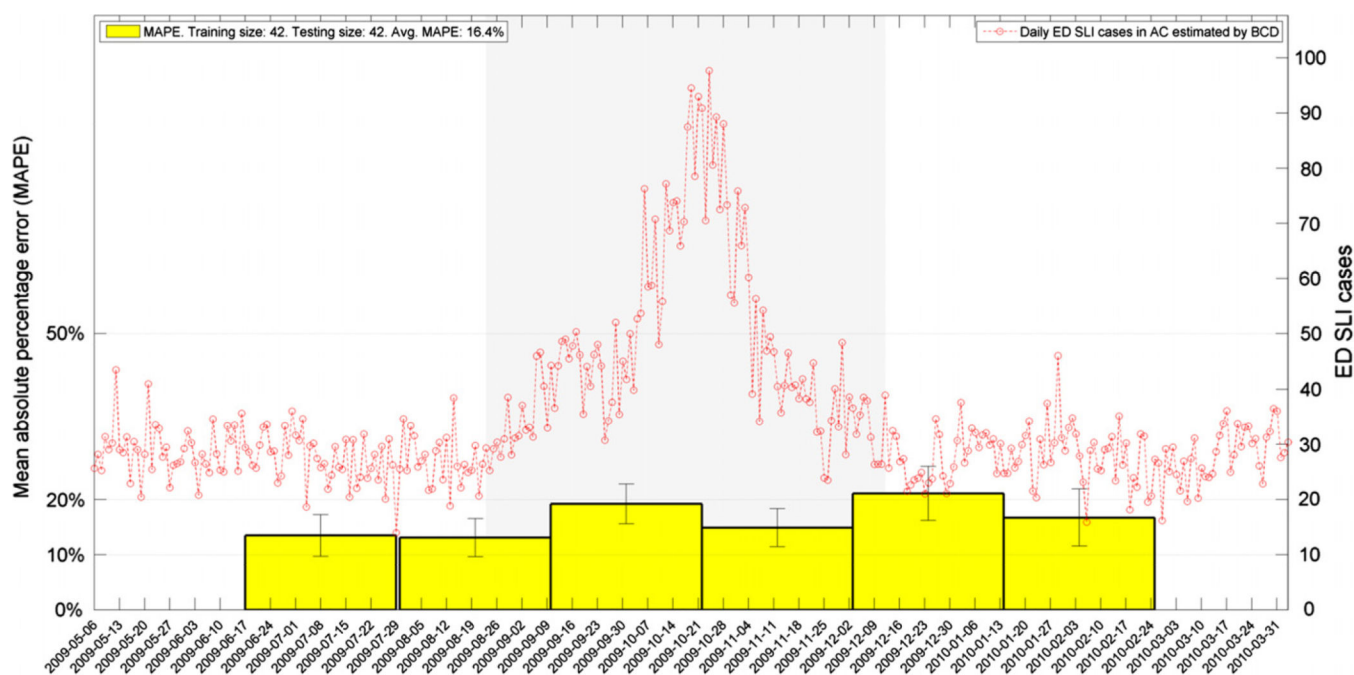


Fig. 4. MAPE values (yellow bars, left y axis) for one configuration in experiment E2 when the training set consists of 42 days (6 weeks) of data and the training set consists of 42 days of data. The error bars represent a 95% C.I. Please see the text and the caption of Fig. 3 for more details about the data shown in this figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Summary of estimates of the incidence of H1N1 cases during the 2009 outbreak.

Authors	Methods	Inputs	Estimates
Reed et al. [1]	Multiplicative model coupled with a Monte Carlo simulation	<ul style="list-style-type: none"> Count of laboratory-confirmed cases through July 23, 2009 Proportions observed in prior published studies and in recent surveys and investigations of pandemic (H1N1) 2009 	Between 1.8 million and 5.7 million H1N1 cases occurred during April-July 2009
CDC [16]	Used Reed et al. method [1]	<ul style="list-style-type: none"> Data on influenza-associated hospitalizations collected through CDC's Emerging Infections Program 	Approximately 55 million people were infected with H1N1 in the US (~18% of its 2009 pop.) between April and December 12, 2009 [16]
Presanis et al. [2]	Bayesian evidence synthesis	<ul style="list-style-type: none"> Cases of medically attended infection in the city of Milwaukee Hospitalizations, intensive care admission or mechanical ventilation, and deaths in New York City Survey data from (1) a New York City telephone survey in May 2009, (2) the Behavioral Risk Factor Surveillance Survey (BRFSS) in nine states in 2007 and (3) a telephone survey in May 2009 in the same nine states as the BRFSS 2007 survey 	From April to July 2009, between 0.16% and 1.44% of symptomatic H1N1 patients were hospitalized and between 0.0007% and 0.048% died
Ross et al. [3]	Hemagglutination-inhibition assays on blood samples	<ul style="list-style-type: none"> Anonymous blood samples obtained from clinical laboratories in Allegheny County, Pennsylvania 	Approximately 63 million people in the US (~21% of the pop.) became infected through early December 2009

Table 2

Datasets used in the experiment E1 and the results obtained. See the text for a discussion of the interpretation of the R^2 and MAPE values.

Number of sets	Training set (days)	Min R^2	Max R^2	Testing set (days)	Avg. MAPE and 95% CI (%)		Min MAPE (%)	Max MAPE (%)
44	21	0.0017	0.6193	7	16.70	±2.38	7.87	46.76
41	42	0.0002	0.7875	7	16.26	±1.99	8.21	39.19
38	63	0.0005	0.8245	7	16.44	±2.14	9.21	38.89
35	84	0.0068	0.8297	7	16.63	±2.43	7.72	38.66
32	105	0.0087	0.8255	7	16.67	±2.49	7.58	34.88

Table 3

Datasets used in the experiment E2 and the results obtained.

Number of sets	Training set (days)	Min R^2	Max R^2	Testing set (days)	Avg. MAPE and 95% CI (%)	Min MAPE (%)	Max MAPE (%)
22	21	0.0017	0.5768	14	16.93 \pm 2.69	9.44	31.12
20	42	0.0009	0.7875	14	16.56 \pm 2.47	9.83	34.57
19	63	0.0061	0.8245	14	16.64 \pm 2.49	9.96	34.08
17	84	0.0103	0.8230	14	17.13 \pm 2.64	11.90	33.90
16	105	0.0300	0.8255	14	16.75 \pm 2.50	9.33	26.99

Table 4

Datasets used in the experiment E3 and the results obtained.

Number of sets	Training set (days)	Min R^2	Max R^2	Testing set (days)	Avg. MAPE and 95% CI (%)	Min MAPE (%)	Max MAPE (%)
14	21	0.0209	0.5768	21	16.98 ±2.87	11.82	28.72
6	42	0.0009	0.7875	42	16.42 ±3.38	13.09	21.12
4	63	0.0061	0.6380	63	19.27 ±3.18	16.53	21.33
2	84	0.0233	0.7862	84	20.17 -	15.34	25.00
2	105	0.0300	0.7484	105	23.75 -	19.23	28.26